DOCUMENT RESUME

ED 239 559 HE 016 983

AUTHOR	Mentkowski, Marcia; Doherty, Austin
TITLE	Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Insights from the Evaluation and Revision Process. Final Report, Research Report Number Three.
INSTITUTION	Alverno Coll., Milwaukee, Wis.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	80
GRANT	NIE-G-77-0058
NOTE	11p.; For related documents, see HE 016 980-990.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Curriculum; College Students; Competence; *Competency Based Education; *Evaluation Criteria; Evaluation Methods; Higher Education; Interrater Reliability; *Liberal Arts; Outcomes of Education; *Research Methodology; *Student Evaluation; Teacher Made Tests; *Test Validity
IDENTIFIERS	*Alverno College WI

ABSTRACT

Perspectives concerning the validation of faculty-developed instruments for the assessment of student performance at Alverno College are presented. Sixteen instruments were identified by departments for the validation studies. Three validation strategies were found to work best. One was a pre- and post-instruction comparison that determined if changes in student performance can be attributed to the effects of instruction. A second strategy was criteria evaluation, which involved the clarification, revision, and refinement of criteria based on an analysis of student performance. A third approach was the interrater reliability of assessor judgments, which enabled a test of reliability as well as the development of instrument criteria. Criteria evaluation appeared to be most helpful when the instrument was being evaluated and revised. Pre- and post-instruction comparisons were used most effectively after faculty had judged the instrument as meeting most other instrument design guidelines. Interrater reliability studies were most useful when they were conducted currently with criteria evaluation. The validation studies showed that direct involvement of faculty in analyzing student performance data and probing validity questions generated a broad scope of validity issues. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED239559

VALIDATING ASSESSMENT TECHNIQUES IN AN OUTCOME–CENTERED LIBERAL ARTS CURRICULUM: INSIGHTS FROM THE EVALUATION AND REVISION PROCESS

Assessment Committee/Office of Research & Evaluation
ALVERNO COLLEGE

FINAL REPORT TO THE NATIONAL INSTITUTE OF EDUCATION:
RESEARCH REPORT NUMBER THREE

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Funded by a grant from the National Institute of Education:
Careering After College: Establishing the Validity of Abilities Learned in College for Later Success
(NIE–G–77–0058)

Principal Investigators:
Marcia Mentkowski
Austin Doherty
Alverno College
3401 South 39th Street
Milwaukee, Wisconsin 53215

NE 016 903

An overview and rationale for our approach to the study of college outcomes, and a summary of the results from the following series of ten research reports, are found in:

Marcia Mentkowski and Austin Doherty. Careering After College: Establishing the Validity of Abilities Learned in College for Later Careering and Professional Performance. Final Report to the National Institute of Education: Overview and Summary. Milwaukee, WI: Alverno Productions, 1983.

Research Reports:

One: Friedman, M., Mentkowski, M., Earley, M., Loacker, G., & Diez, M. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Valuing and Communications Generic Instrument, 1980.

Two: Friedman, M., Mentkowski, M., Deutsch, B., Shovar, M.N., & Allen, Z. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Social Interaction Generic Instrument, 1982.

Three: Assessment Committee/Office of Research and Evaluation. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Insights From the Evaluation and Revision Process, 1980.

Four: Assessment Committee/Office of Research and Evaluation. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Integrated Competence Seminar, 1982.

Five: Assessment Committee/Office of Research and Evaluation. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Six Performance Characteristics Rating, 1983.

Six: Mentkowski, M., & Strait, M. A Longitudinal Study of Student Change in Cognitive Development and Generic Abilities in an Outcome-Centered Liberal Arts Curriculum, 1983.

Seven: Much, N., & Mentkowski, M. Student Perspectives on Liberal Learning at Alverno College: Justifying Learning as Relevant to Performance in Personal and Professional Roles, 1982.

Eight: Mentkowski, M., Much, N., & Giencke-Holl, L. Careering After College: Perspectives on Lifelong Learning and Career Development, 1983.

Nine: Mentkowski, M., DeBack, V., Bishop, J., Allen, Z., & Blanton, B. Developing a Professional Competence Model for Nursing Education, 1980.

Ten: Mentkowski, M., O'Brien, K., McEachern, W., & Fowler, D. Developing a Professional Competence Model for Management Education, 1982.

© Copyright 1980. Alverno College Productions, Milwaukee, Wisconsin. All rights reserved under U.S., International and Universal Copyright Conventions. Reproduction in part or whole by any method is prohibited by law.

ABSTRACT

The Alverno College faculty has designed a curriculum and assessment process to assist students to develop and demonstrate ability in a variety of competences. Faculty, individually and as a group, design assessment instruments which then come under the scrutiny of other faculty in a continuous process of review and redefinition. This evaluation and revision process stimulates evaluation and revision of the instruments in a systematic way.

Validating assessment instruments is an unusual goal for a college faculty to pursue. To validate means that concepts of the abilities or competences assessed and the means for doing so must be carefully thought out, subjected to rigorous reasoning, and constantly reviewed against student performance outcomes. This report summarizes questions, suggestions, concerns and insights generated from feedback sessions with faculty who submitted their instruments for a validation study. Sixteen instruments were identified by departments as ready to submit because faculty judged them sufficiently developed to evaluate. Three validation strategies worked best of those tried. One is pre- and post-instruction comparison which determines if changes in student performance can be attributed to the effects of instruction. A second is criteria evaluation, which involved the clarification, revision and refinement of criteria based on an analysis of student performance. A third is establishing the inter-rater reliability of assessor judgments, which enables a test of reliability as well as the development of instrument criteria. Criteria evaluation appears to be most helpful when the instrument is being evaluated and revised. Pre- and post-instruction comparisons are used most effectively after faculty have judged the instrument as meeting most other instrument design guidelines. Inter-rater reliability studies are most useful when they are conducted concurrently with criteria evaluation. The validation studies that were synthesized for this report show that direct involvement of faculty in analyzing student performance data and probing validity questions generates a broad scope of validity issues.

ACKNOWLEDGEMENTS

We wish to acknowledge the contributions of the following persons:

Assessment Committee (1977–82):

Zita Allen
Jean Birkey
Austin Doherty
Margaret Earley
Miriam Friedman
George Gurria
Pat Jensen
Joan Koehler
Cathleen Krzyminski
Dimitri Lazo
Georgine Loacker
William McEachern
Marcia Mentkowski
Agnes Meysenburg
Marlene Neises
James Newton
Kathleen O'Brien
James Roth
Jean Schafer
Judeen Schulte
Kyle Stewart
Michael Strait
Alice Theine
Christine Trimberger
Allen Wutzdorff

VALIDATING ASSESSMENT TECHNIQUES IN AN OUTCOME-CENTERED LIBERAL ARTS CURRICULUM: INSIGHTS FROM THE EVALUATION AND REVISION PROCESS

Assessment Committee/Office of Research and Evaluation [1]

Introduction:

Validating assessment instruments is an unusual goal for a college to pursue. Historically, each college professor has developed most of his or her own instruments to assess student performance. An individual professor might employ various methods to improve testing instruments--but seldom if ever would he or she submit them to others for systematic and continuous review. Nor are systematic attempts often made to compare a student's performance across a number of courses or instruments, or to predict future professional success from measures of student performance in college courses.

The Alverno College faculty has set itself the task of assisting students to develop and demonstrate ability in a variety of competence areas (e.g. Communications, Analysis, Problem Solving, Valuing in Decision-Making, Social Interaction, etc.) that faculty as a group have chosen through consensus as important to individual growth and professional performance. They have implemented an assessment system where faculty, individually and in groups, design assessment instruments which then come under the scrutiny of others in a continuous process of review and redefinition. Quality assurance procedures stimulate evaluation and revision of instruments in a systematic way.

Because faculty extended themselves to collaborate on assessment, as well as on specifying goals, they are necessarily asking further questions about the validity of their instruments or techniques.

What do faculty mean by validating the techniques of assessment? In this liberal arts college, "to validate" means that concepts of the abilities assessed and the means for doing so must be carefully thought out, subjected to rigorous reasoning, and constantly reviewed. Among the immediate responses to this commitment has been reliance on the objective judgment available from a variety of sources--faculty judgment across disciplines and competences, judgments from professionals outside the institution who also serve as assessors of student performance, and special interdisciplinary committees like the Assessment Committee set up to generate objective judgments about the individual assessment techniques and to monitor instrument evaluation procedures. Faculty have committed themselves to go beyond content validity and evaluation and revision of instruments, to questions of validity because their questions encourage it.

Faculty Conducted Validation Studies: Some Insights

The following report summarizes questions, suggestions, concerns and insights generated from feedback sessions with faculty who submitted their instruments for a validation study. Sixteen instruments were identified by departments as "model" instruments because faculty judged them sufficiently developed to validate.

Several validation strategies were employed by faculty:

[1] This paper was reproduced in Robert V. Lorenzo (Ed.) Proceedings of the Eighth International Congress on the Assessment Center Method; Toronto, Ontario, June 4-6, 1980

Pre- and Post-Instruction Comparison

This procedure provides information on the extent to which the instructional process produces changes in students' performance and the extent to which the instrument under study is effective in measuring such changes.

Criteria Evaluation

Clarification, revision and refinement of criteria based on an analysis of student performance brings us closer to the intended meaning of the behaviors and abilities measured, thereby creating a more valid assessment technique.

Inter-rater Reliability of Assessor Judgments

In our work with "model" instruments that assess generic, developmental and holistic competences (Alverno College Faculty, Assessment at Alverno College, 1970) we are often inferring an unobservable "construct" or ability from observed behavior. It is essential to continue to develop our understanding of the nature of the competence or ability we teach toward "construct validity" by integrating evidence from different sources of expert judgment. Establishing inter-rater reliability of judgments by two or more assessors remains one of the better ways to establish instrument validity. Comparing our professional judgments stimulates development of mutual standards as a base for defining instrument criteria.

The model instrument validation studies demonstrate that direct involvement of faculty in analyzing student performance data and probing validity questions generates a broad scope of validity issues.

An important outcome of some feedback sessions with faculty was the recognition that many kinds of "validations" will result in a "qualitative" rather than a quantitative analysis. For example, one faculty member, after comparing each student on a general pre- and post-assessment based on classroom observations, was able to identify the number of students who had gained more of the objectives, some of the objectives, and few if any of the objectives. Still another criteria evaluation was completed by one member of the Assessment Committee who simply counted the number of students who completed each of the objectives, based on data collected by the instructor during the semester--data the instructor used to record information for individual feedback and competence validations.

Still another important outcome of the feedback sessions, and the information relating to validating the model instruments that we have collected so far, is that most criteria evaluations will not involve collecting more data than we already collect in our role as a course "assessor." In general, we learned that not all three validation strategies need to be employed concurrently. Criteria evaluation appears to be most helpful when the instrument is being evaluated and revised. Pre- and Post-Instruction comparisons are most helpful when faculty have judged the instrument as generally satisfactory. Inter-rater reliability studies are probably most useful when they are used concurrently with criteria evaluation.

Insights from Pre- and Post-Instruction Comparisons

In conducting a Pre- or Post-Instruction study establishing the reliability and the validity of an instrument and the assessment outcomes, we must consider the composition of the student group involved. One group may be a homogeneous group in that students have similar areas of concentration, developmental stage, motivation,

or purpose for pursuing the course. Another group may be more heterogenous with respect to these factors. Students may be from a variety of majors, their year in school may be different, some students may be taking the course primarily to meet certain validation requirements, etc. Such homogeneity or diversity may very well be reflected in the kinds of learning experiences they choose after the pre-assessment, the nature of the post-assessment, and the expectations for validation. These variations may need to be considered in the overall interpretation of the student performance data.

Another factor to be considered is the degree to which students are motivated to perform. It is important to create a comparable motivational effort in both the pre- and post-assessment. A powerful motivation for pre-assessment is "testing out" of a competence level. If students are told to regard the pre-assessment administration as only a source of information to the instructor, and there is no tangible benefit to them personally, lack of motivation alone may account for differences between the pre-and post-assessment.

We also found a need to look at the relationship between the pre- and post-assessment. Are they comparable with respect to instrument stimulus, to criteria, to mode of assessment? In some disciplines, where competence is inseparable from content, it may be impossible to administer the same stimulus. Students are not yet familiar enough with the complexity of the content during a pre-assessment. In that case, instructors may decide to administer two different stimuli. But will student performance be comparable, even though it is evaluated by the same set of criteria? This issue points out the extent to which student performance may vary with the stimulus employed.

Another insight that emerged cautioned us to examine the "route of progress" as well as the "rate of progress" in comparing performance from a pre- to post-assessment. The instrument is powerful diagnostically if progress can be qualitatively evaluated rather than providing a statement of all or none progress alone.

Insights from Criteria Evaluations

Several questions have emerged as we have discussed criteria evaluations with faculty.

First, we have found that it is important to ask whether students have enough opportunity to demonstrate the called-for behaviors. For example, an assessment technique which asks students to demonstrate a number of competences may provide less opportunity for the student to be explicit in responding to one competence.

Second, we found that lack of mastery of one competence may impact student performance in a related competence. For example, students who did not master levels 1 and 2 in Analysis may not reach an accepted level of performance in Communications, Listening and Reading, levels 1, 2, 3 and 4.

Another question we must ask is: Will instruments tend to be more "valid" for assessing the effects of instruction if criteria are presented to the student explicitly? How does this consideration affect criteria from levels 5 and 6, which are deliberately more implicit?

Finally, what is the relationship between criteria and content? Are there content areas which are more readily integrated with a specific competence? How do criteria change along with the content of the discipline and situational variations in administering the instrument?

Insights from Studies of Inter-rater Reliability of Assessor Judgments

It seems important to further investigate the inter-rater reliability of an instrument by giving instruments to practicing professionals off-campus in order to learn how they interpret student performance, as compared to the educational expert on-campus. Such a procedure may provide an additional measure of the external validity of an instrument.

Some faculty members were particularly interested in re-judging student performance already judged by another instructor (study of the inter-rater reliability of assessor judgments) because the faculty member wanted to know how close he/she would come to understanding the criteria in the same way. This faculty member was interested in stimulating additional discussion about the criteria currently under study in the department.

In the absence of judgments from two assessors, is it possible to make a prediction as to the consistency with which student performance on the instrument might be judged by another faculty member from the same discipline? Some individually designed instruments used within a specific course as a formative assessment may be so specific that it may be difficult to find another assessor to make a second judgment.

Some faculty who conducted a study of the reliability of assessor judgments began thinking about level 5 and 6 validations and the role of criteria in eliciting student performance. They asked: What is the relationship of assessors' judgments to explicit vs. implicit criteria? Will judgments be consistent with the same set of criteria when one is defined in an open-ended way and the other is more directive? For example, can one assessor's judgment be expected to be similar to the judgment of another assessor if the criteria are defined explicitly vs. implicitly? Supposedly, if the criteria are directive, they may elicit performance that is different from criteria that are implicit in the instrument directions to students.

In conclusion, this first group of faculty-conducted validation studies provided important insights for future work. At each step of the way, we have come to recognize the importance of group effort in pursuing validation issues. Insights from one department assist another. More important, we are finding instrument validation easier in some ways than we originally thought. Faculty have a great deal of experience individually in instrument "validation"--even though they may call it something else in their own mind. Sharing ideas in these feedback sessions with faculty has clarified our thinking and has supported our efforts to continue to pursue validation issues with our non-traditional assessment techniques.

CRITERIA FOR ASSESSMENT INSTRUMENTS

1. Does the instrument elicit and measure the complex abilities designated for competence level(s) within a defined context?

2. Does the instrument elicit the fullest expression of student ability at that level in that context?

3. Does the instrument require the use of substantive content commensurate with the level of sophistication of the ability?

4. Does the instrument integrate previous levels of the competence and require the student to demonstrate an increasingly sophisticated ability of lower levels at higher levels?

5. Does the instrument elicit a range of performance?

6. On a scale from discrete to fully integrated, does the instrument reflect the appropriate level of integration of dimensions of performance (content with competence; among competences)?

7. Does the instrument involve a production task rather than a recognition task?

8. Does the instrument use an assessment mode that recognizes the intrinsic nature of the ability being assessed?

9. Does the instrument allow for the judgment of performance against public and explicit criteria?

10. Does the instrument assess the student's ability to self-assess?

11. Does the instrument allow for assessment of the student's performance external to the learning situation?

12. Does the instrument elicit performance with sufficient data to provide for diagnostic, structured feedback to the student on her strengths and weaknesses?

13. Do the instrument criteria provide evidence for credentialing performance?

0581

Alverno College
3401 South 39th Street / Milwaukee, WI 53215